A High Fidelity Multi-Sensor Scene Understanding System for Autonomous Navigation

Mark Rosenblum
Center for Intelligent Systems
Science Applications International Corp.
Littleton, CO 80127
rosey@cis.saic.com

Benny Gothard
Center for Intelligent Systems
Science Applications International Corp.
Littleton, CO 80127
bgothard@cis.saic.com

Abstract

In order for an autonomous military robot to "appropriately" navigate through a complex environment, it must have an in-depth understanding of the immediate surroundings. In the military sense, appropriate navigation implies the robot will avoid collision or contact with hazards, will not be falsely re-routed around traversible terrain due to false hazard detections, and will exploit the terrain to maximize its concealment. Appropriate autonomous navigation requires the ability to detect and localize critical features in the environment in order to respond appropriately to them. We have developed a scene understanding system based on a multi-sensor system that uses an "operator-trained" rulebase to analyze the pixel level attributes across the set of diverse phenomenology imaging sensors. Each pixel is registered to range information so we not only know what but where features are in the environment. This three dimensional labeled world model can then be used to control the speed and steering of the vehicle in an appropriate manner. In this paper we discuss our multi-sensor system, the operator trained analysis algorithm called ONAV (Opportunistic NAVigation), and the reactive control algorithm used to control the speed and steering of the vehicle.

1. Introduction

A great deal of technology has to be developed in order to achieve the goal of developing a fully autonomous vehicle. One of the main "tall poles" that must be overcome in order for this goal to come to fruition is the ability of the computer to understand its surrounding environment to a level that is required for the intended task. The military mission scenario requires a robot to interact in a complex, unstructured, dynamic environment.

2. Mobility Requirements

The military scenario target operating environment is the battlefield. A robotic vehicle must function in daytime and nighttime and in a variety of reduced visibility conditions such as fog, dust, smoke and airborne precipitation. This operating environment imposes challenging requirements on the computer vision system especially since the system must function with a high degree of reliability, robustness and safety in this diversity. Changes in operating conditions can be immediate or gradual. Immediate changes can occur because the robot is moving and interacting in the environment and can encounter condition transitions such as moving from direct sunlight to shade. The movement of the robot in the environment also causes the robot to see features from different aspect views. For instance, the robot must be able to identify a vehicle whether it is viewing it from the front, side or rear. Gradual condition and environment changes occur at different time scales. In the course of a day, the lighting can change because clouds block the sun or the daytime transitions to nighttime. In the seasonal cycle of a year, sun angles vary and the environment can take on a completely different look from heavy green foliage to leafless trees and brown grass. With all of these challenges in mind, one must not forget that the main purpose of the computer vision system is to visually servo the robot in the environment for safe autonomous control that achieves human level control performance. This imposes a real-time requirement on the computer vision system. The speed requirements for a robotic military vehicle are 20mph off-road and 40mph on-road in daylight, and 10mph off-road and 20mph on-road at nighttime or

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A High Fidelity Multi-Sensor Scene Understanding System for Autonomous	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Science Applications International Corp,Center for Intelligent Systems,Littleton,CO,80127**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**7**	

degraded visibility conditions. This translates to a cycle rate of 10Hz for 20mph and 20-25Hz for 40mph [1]. The higher cycle rates for 40mph on-road operation will be achieved by focusing the processing on a restricted set of environmental features augmented with the use of obstacle detection radars to detect obstructions in the roadway.

3. Approach

We have applied a systems philosophy to the computer vision problem, and we have designed a system called ONAV (Opportunistic NAVigation) that can harness all of the computer vision technology to date, and combine these approaches into one integrated system. In ONAV, no one sub-component bears the burden of the problem. In other words, we do not expect algorithms alone to solve the computer vision problem. If we choose effective sensing that inherently performs some level of scene discrimination at the phenomenology level, algorithms will be handed a partially analyzed scene before they ever encounter the raw image data. The algorithms have been designed to exploit an optimized processing hardware infrastructure, to maximize computation for the "real-time" application of autonomous robot navigation.

ONAV will: 1) identify hazards in the scene to which the robot must avoid, 2) identify beneficial features with which the robot must interact, and 3) provide data for route planning over a 50-1000 meter planning horizon. The application of computer vision to autonomous robot navigation has several key characteristics that set it apart from other computer vision applications, and we have designed ONAV based on these characteristics. Since ONAV is constantly running during a robot traversal of the terrain, it will have multiple looks, and thus multiple attempts to analyze approximately the same scene. In addition, since vehicle response is typically occurring while the vehicle is moving, the scenes presented to ONAV are gradually changing. This provides ONAV with different perspectives on the same features in the scene, and different chances on slightly different input to make a correct interpretation. This also allows a scene to be analyzed over a sequence of images, increasing the accuracy and certainty of an interpretation. This lowers the level of the single frame accuracy requirement for the autonomous navigation vision system.

In order to interpret a scene for a particular application, it is necessary to have distinct classification categories. What makes a classification problem difficult is the similarity of different categories or lack of separability in attribute space of these categories. The classification categories for autonomous navigation are typically wide-ranging and have high degrees of separability between categories. The categories are wide-ranging for a particular feature type due to the fact that fine resolution of categorization is not required. For instance, in classifying grass, it is not necessary to tell the difference between Kentucky Blue Grass and Fescue, or in classifying a rock it is not necessary to discriminate between granite and limestone, only that it is grass or a rock, respectively. In both of these classifications, it is only necessary to classify at a gross level of resolution.

Generally, separability of features in the environment occurs naturally for the feature set required for autonomous navigation. Typically, different features in the environment have different characteristics or else they would be considered the same feature. For this reason, humans can identify and appropriately interact with features in the environment. One of the key philosophies of our design is that we attempt to utilize the same visual cues as a human driver but for an extended set of phenomenologies. For instance a hazardous rock in view in the grass will have separability from the grass through the attributes of elevation, spectral properties, texture, shape, and thermal properties. Even one of the more traditionally difficult hazardous features to detect, a negative obstacle, which is a hole or rut in the terrain, shows up as intensity and texture discontinuities in visible imagery and as texture and thermal discontinuities in thermal imagery.

4. ONAV System Description

ONAV is a culmination of the systems philosophy applied to the computer vision problem for autonomous robot navigation. It involves algorithms, sensors and processing architecture as depicted in Figure (1). The purpose of ONAV is to provide an infrastructure to combine multiple sensing modalities, visual cues, and algorithms into

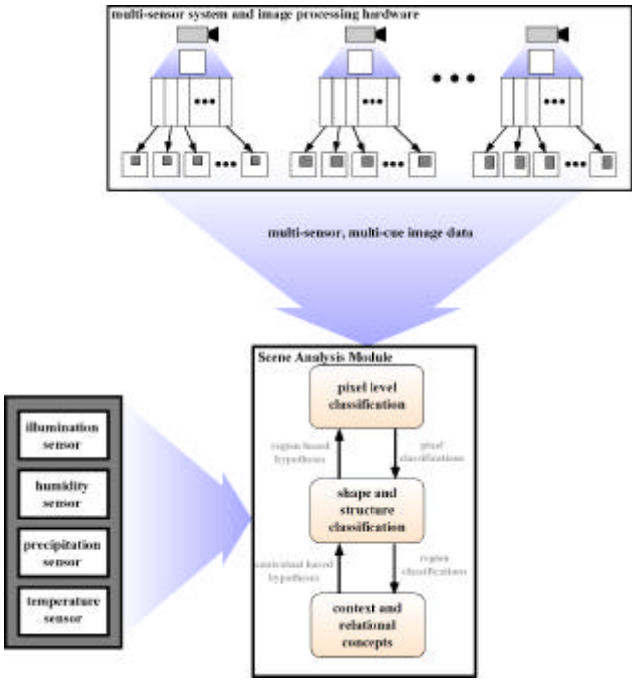

Figure (1): ONAV System

one cohesive system to robustly interpret a scene for the application of robot navigation. In this section, we describe the sensor system, processing architecture and algorithms that makeup the ONAV system.

4.1 Sensor System

Most imaging sensors do not have the same wide operating ranges as the human visual system. In order to achieve human level operating ranges and beyond, the ONAV sensor system was designed using a two pronged approach:

1) optimize the data from each individual sensor using environmental sensing,
2) integrate diverse sensor phenomenologies sensors with independent operating ranges into one "system" so when combined the individual operating ranges cover the entire extent of the target operating range for the application.

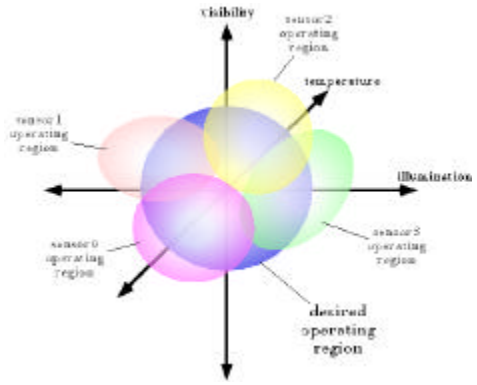

Figure (2): The overlapping of diverse sensor operating conditions can achieve the desired operation conditions.

Figure (2) shows how the union of individual sensor operating ranges can achieve the required operating range over all conditions for the application. Some of the key environmental conditions that can affect system performance are illumination, humidity, airborne precipitation, smoke, haze and fog. By selecting a set of sensors such that there is always a subset that is minimally impacted by extreme operating condition, useful information about the environment will always be available for analysis. For example, if a FLIR is combined with a color camera, and a range imaging device such as a scanning laser range finder, there will almost always be a sensor that works in daylight, nighttime, precipitation, fog, smoke and haze. In addition to the imaging sensors in the ONAV system, it would also be advantageous to exploit non-imaging sensors such as RADARS or ultrasonics for hazard detection.

4.2 Processing Architecture

In addition to using raw sensor data from multiple sensors, ONAV also makes use of low-level visual cues such as texture, edges and range. This presents multiple problems. First, many of these low-level processing cues require extensive computation. Second, each low-level processed visual cue is a data band in itself such as the red band in a color

image, and the additional set of processed data bands, like the raw data bands directly from the sensors, must be accessible by the ONAV software task. This presents issues for data flow and memory resources. Third, our processing architecture design for ONAV is not just for ONAV, but for all of the other processes in the system such as road-following, path-following, vehicle tracking, reconnaissance, surveillance and target acquisition. With these key issues in mind, the processing architecture to support ONAV and the other processes in the system resulted in a distributed processing architecture with the sensors tied as closely to the tasks in the architecture that need them. The computationally intensive low-level visual cue processing tasks are allocated to individual dedicated processors so this data is available in a real-time manner. Where ever possible, there was an attempt minimize bus traffic and utilize auxiliary buses independent of the main bus structure to pass data.

4.3 Algorithm

The ONAV software consists of many different components that will reside on different processors in the system. Figure (3) shows the hierarchy of software in the ONAV system.

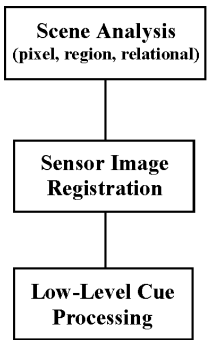

Figure (3): Software Hierarchy in the ONAV system.

The lower level layers of the software hierarchy, low-level cue processing and image registration, provide the supporting elements to perform the scene analysis which takes place in the pixel level processing layer and up. In this section we will summarize the development that has occurred at each level in the software hierarchy.

4.3.1 Low-Level Visual Cues

It is known that the human visual system makes use of multiple visual cues in the analysis of a scene. The importance of this multi-visual cue set is illustrated by the fact that a one eyed person can successfully navigate and recognize features in his/her environment. With vision in only one eye, this person loses depth perception based on the visual cue of binocular disparity. To compensate for this loss, this person enhances the importance of other visual cues that can be gleaned from a scene such as contrast, color, texture, perspective, shape, size, motion, and orientation. ONAV utilizes a similar set of visual cues to that used by the human visual system. The use of multiple visual cues increases the number of discriminants in the classification of features in the scene and also improves the reliability and robustness of the analysis. Since many visual cues will be incorporated into the analysis, a lapse in one cue will only cause a slight degradation in the performance in the overall system performance, where a single cue system would abruptly fail. Additionally, some of the low-level visual cues are naturally semi-invariant to changing conditions in the environment, which improves analysis performance of the system over a wider range of operating conditions.

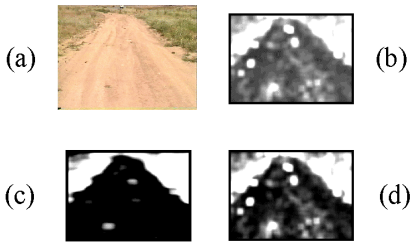

Figure (4). Homegrown Texture approach a) road bounded by grass, b) entropy, c) horizontal transitions, and d) vertical transitions.

4.3.1.1 Texture

It has been shown in biological vision that texture is one of the most powerful visual discriminants, however, the biological system almost never relies on texture alone to interpret a scene. ONAV will also not rely solely on texture to analyze a scene, but only use it to improve the analysis. We have evaluated the following texture analysis techniques for inclusion into the ONAV system: Co-Occurrence

Matrices [2], Gabor Filters [3], Markov Models [4], and some simple "home-grown" techniques. We found the simple home-grown techniques worked best for our application. Figure (4) shows some sample analyses using the homegrown texture approach.

4.3.1.2 Stereo Range

The role of the passive stereo vision system is to provide a dense set of 3D measurements in front of the vehicle. We are using an implementation of area-based stereo matching provided by the Jet Propulsion Laboratory [5]. In this implementation, the disparity between left and right image at a given pixel is found by searching along the epipolar line for the pixel with the highest correlation of intensity values in a window centered at that pixel. The local windows are compared using the Sum of Squared Differences (SSD). Efficient search is achieved by first rectifying the images such that the epipolar lines are parallel to the image scanlines. We are currently using an outdated version of a stereo implementation which runs on a DataCube MV200 at a 2 Hz rate for a 200x70 subimage. It is not uncommon for current implementations of stereo on off-the-shelf hardware to cycle at 10 Hz. Active LADAR can also provide this form of data.

4.3.1.2 Specialized Feature Detectors

Certain obstacles such as negative depressions are difficult to detect with range data alone, and by only focusing on the range data, a significant amount of discriminating information is thrown away that could contribute to the detection of these hazards. To improve the detection robustness for such obstacles, we have implemented specialized feature detectors that will work on any form of intensity image, be it visible spectrum or thermal. Examples of a specialized feature detector are a horizontal or vertical band detector. When using a horizontal band detector to help detect negative obstacles, the band dimensions need to change with distance from the vehicle due to the effects of perspective. Essentially, the band detectors are "sandwich" difference operators where the absolute difference between the average of the pixel intensities in the inner and outer window produce the response from the operator. Figure (5) shows the response of the horizontal and vertical band detectors.

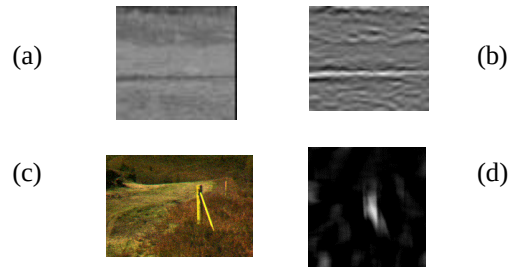

Figure (5): a) intensity image of negative obstacle, b) horizontal band operator response, c) fence post, d) vertical band operator response.

4.3.2 Multi-Sensor Image Registration

In order to make use of multiple imaging sensors for fusion at the pixel level, it necessary that all of the imaging sensors be registered. Since all of our sensors have their own optics, and their imaging planes have physical offsets from each other, it is impossible to obtain an exact registration solution. For our application we have built an approximate registration model that will have accuracy down to several pixels, which in most cases will be sufficient. The approximate mathematical registration model is represented by a quadratic polynomial with 6 coefficients [6]. The model is adjusted by adjusting the 6 coefficients. In this registration scheme, one sensor is chosen as the reference sensor, and all pixel locations in the classifier are within this reference sensor coordinate system. Each sensor in the system requires a registration model to the reference sensor so pixels from each non-reference can be transformed into the reference sensor coordinate system.

4.3.3 Fusion of Data

Computer Vision algorithms typically have to deal with large amounts of data since a typical color image of dimensions 640x480 consists of 900K bytes. In ONAV where we are dealing with multiple sensors and multiple visual cues, the issue of dealing with large amounts of data is even more exaggerated, and a fusion method is required to intelligently combine all of this data to produce useful results. We have developed a fuzzy logic

rulebase system that encapsulates visual attributes into rules through a process of "human-guided-training". Our approach was borrowed from the techniques used to train automated satellite image analysis systems [6]. The human trained classification system consists of three layers of analysis: 1) pixel, 2) region, and 3) relational or dependencies between feature types. The learning at all layers is supervisory in nature and the supervised learning signal is the identification of terrain features and types in the scene by the human trainer. With this learning model, as the system is exposed to more sites and conditions, it will be able to classify more terrain without compromising previously learned information. Figure (6) shows pixel analysis results for a typical scene using three bands of color data, one band of thermal data, stereo range data, and the low-level processed texture cue. To this point most of our effort has been on pixel level analysis, and we have only just begun to exam region and relational analysis.

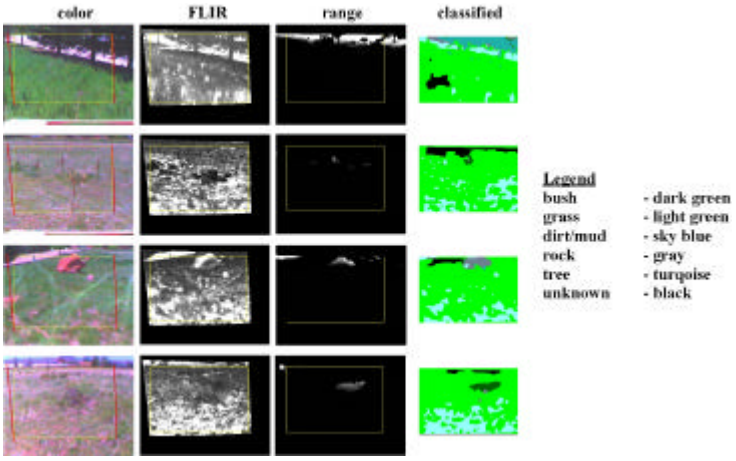

Figure (6). Pixel Level Classification Results combining color, thermal and range data. Row 1: trees and shadows, Row 2: tall grass, Row 3: rock in grass, Row 4: bush in grass.

4.3.3.1 Achieving Analysis Robustness

All models have a target range of operability. This is also true for the fuzzy logic rulebases built through the human guided training process described above. We have taken a somewhat exhaustive approach to covering the full range of operating conditions that ONAV must handle. Instead of putting the burden on one rulebase to handle all sets of conditions, we will have many rulebases, each tuned to a unique range of operability, and collectively, the set of rulebases will cover the entire extent of target operating conditions.

With all of these rulebases for a particular terrain type, some sort of high-level coordination is required to manage the system and optimize performance based on the current set of conditions and environmental characteristics. For this we have designed a Meta-Rulebase-Manager (MRM). The MRM has several responsibilities:

- monitor the performance of the individual rulebases and activate the most applicable to the current set of conditions
- monitor sensor performance and phase-out or phase-in sensors as they become operable in the current set of conditions.

4.3.3.2 Complete Classification Failure

We expect that the classification system will not be perfect. We are expecting mis-classifications, and the possibility of "null" classifications or unknown classification response. Our system has inherent fallbacks built directly into the architecture. If ONAV is unable to make a classification, the fallback will be to use just 3D range information to navigate the vehicle. The fallback to poor stereo range data will be to use the LADAR and radar system. With all of these layers of sensor phenomenologies, the system will be robust, reliable and safe in a wide range of operating conditions.

5. Navigating with Classification

Once the classified image is generated, it can then be used to appropriately control the vehicle. For our military scenario, appropriate navigation requires maximizing vehicle concealment and minimizing mission time, while keeping the vehicle safe at all times. We have designed and built a complementary module called OREACT (Opportunistic Reaction) that acts as the reactive control backend to ONAV. OREACT is reactive in the sense that it is responding to the immediate surrounding without a notion of a mission goal. OREACT takes in the classified image from ONAV, enters the labeled imagery into a discretized three dimensional world model using the registered range information tagged to each labeled pixel, and evaluates potential steering directions and speeds against the appropriate mobility requires. The control response from OREACT is a vector representing the scores of the potential steering directions. Because the response from OREACT is not a single steering and speed response, but a set of valid responses, the response vector from OREACT must be superimposed with the response vector from an additional behavior that provides a specific motivating steering and speed response such as a way point follower. Figure (9) shows the three-dimensional world model built by OREACT with the labeled data from ONAV along with the viable steering arcs whose arc lengths are proportional to the computed safe speed response.

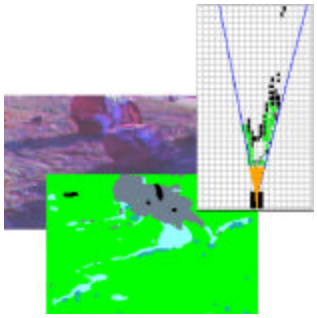

Figure (9): OREACT three-dimensional World Model used to generate reactive steering and speed.

6. Results

Our combined scene analyzer, ONAV, and our reactive planner, OREACT, robustly avoid hazards, runs over traversible features such as small bushes and tall grass, and hugs features such as trees, tall bushes and rocks that will conceal the vehicle from an overhead and a ground perspective. The integrated system works in a wide variety of lighting conditions from bright sunlight with lots of shadows to overcast conditions. With the appropriate sensors such as a stereo FLIR system, we will be able to extend a subset of this capability to night time operation. Our current cycle time is about 2 seconds, which is much to slow, but we are planning on upgrading our processing infrastructure and expect a cycle time on the order of at least 10-20 Hz.

Acknowledgements

This work was supported by OUSD(A&T) Joint Robotics Program, TARDEC under contract number DAAE04-98-C-L013 and the DARPA MARS Program under contract number DABT63-00-C-1008.

References

[1] M. Rosenblum, B. Gothard, B. Klarquist, J. Kurtz, “Autonomous Mobility for Demo III,”, SPIE International Symposium on Intelligent Systems and Advanced Manufacturing: Mobile Robots XIII, Paper 3525A-2, 1998.

[2] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, vol. 1, Addison-Wesley, Reading, MA, 1992.

[3] H. Greenspan, “Non-Parametric Texture Learning,” in *Early Visual Learning*, Editors S.K. Nayar, T. Poggio, Oxford University Press, pp. 299-328, 1996.

[4] R. Chellappa, S. Chatterjee, “Classification of Textures Using Gaussian Markov Random Fields,” in IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-33, no. 4, August 1985.

[5] M. Hebert, R. Bolles, B. Gothard, L. Matthies, M. Rosenblum, “Mobility for Unmanned Ground Vehicles,” in *Reconnaissance, Surveillance, and Target Acquisition for the Unmanned Ground Vehicle: Providing Surveillance ‘Eyes’ for an Autonomous Vehicle*, Editors O. Firschein and T.M. Strat, Morgan Kaufmann, pp. 95-108, 1997.

[6] R.A. Schowengerdt, *Remote-Sensing: Models and Methods for Image Processing*, Academic Press, San Diego, CA, 1997.